HIERARCHICAL MODELS, MARGINAL POLYTOPES, AND LINEAR CODES

THOMAS KAHLE, WALTER WENZEL, AND NIHAT AY

Abstract. In this paper, we explore a connection between binary hierarchical models, their marginal polytopes and codeword polytopes, the convex hulls of linear codes. The class of linear codes that are realizable by hierarchical models is determined. We classify all full dimensional polytopes with the property that their vertices form a linear code and give an algorithm that determines them.

1. Introduction

In theoretical statistics the marginal polytope plays an important role. It is the polytope of possible values that a sufficient statistics can take. It encodes in its face lattice the combinatorial structure of the boundary of the exponential family defined by the statistics. For a model on discrete random variables it can be represented with vertices that have only components 0 or 1, commonly called a 0/1 polytope.

In coding theory when decoding binary linear codes one can apply techniques from linear programming and optimize a linear function over the convex hull of the code words, known as the codeword polytope [FWK03].

Observing that for certain choices of sufficient statistics on binary random variables these two notions coincide, our main contribution is a characterization of the corresponding polytopes. We do not address problems that are directly linked to coding theory. However, we do hope that our result will contribute to a better understanding of the closure of exponential families, which is an important problem in statistics.

The paper is organized as follows: In Section 2, we introduce the necessary notions to define hierarchical models and fix the notation. We review different descriptions of so called interaction spaces in Section 3. In Section 4, we establish the link to coding theory. Finally, in Section 5 we give our main result, the classification of all such full dimensional polytopes whose vertices form a linear code and give a recursive formula for their number.

2. Preliminaries

2.1. **Exponential Families of Hierarchical Models.** Given a non-empty finite set $\mathcal{X}$, we denote the set of probability distributions on $\mathcal{X}$ by $\overline{\mathcal{P}}(\mathcal{X})$. The *support* of $p \in \overline{\mathcal{P}}(\mathcal{X})$ is defined as $\mathrm{supp}(p) := \{x \in \mathcal{X} : p(x) > 0\}$. The set of distributions with

full support is denoted as $\mathcal{P}(\mathcal{X})$. The set $\overline{\mathcal{P}}(\mathcal{X})$ has the geometrical structure of a $(|\mathcal{X}| - 1)$-dimensional simplex lying in an affine hyperplane of

$$\mathbb{R}^{\mathcal{X}} := \{f : \mathcal{X} \to \mathbb{R}\},$$

the vector space of real valued functions on $\mathcal{X}$. Statistical models, such as hierarchical models are subsets of $\overline{\mathcal{P}}(\mathcal{X})$. In this paper, we will only consider so called exponential families which are smooth manifolds.

Definition 1. The map

$$\exp : \ \mathbb{R}^{\mathcal{X}} \to \mathcal{P}(\mathcal{X}), \qquad f \mapsto \frac{\mathrm{e}^f}{\sum_{x \in \mathcal{X}} \mathrm{e}^{f(x)}},$$

is called the *exponential map*. It acts component wise by exponentiating and normalizing. Then, an *exponential family (in $\mathcal{P}(\mathcal{X})$)* is defined as the image $\exp(\mathcal{I})$ of a linear subspace $\mathcal{I}$ of $\mathbb{R}^{\mathcal{X}}$.

An exponential family $\mathcal{E}$ naturally has full support and is therefore contained in the open simplex $\mathcal{P}(\mathcal{X})$. However, to get probability distributions with reduced support one has to pass to the closure $\overline{\mathcal{E}}$ with respect to the standard topology of $\mathbb{R}^{\mathcal{X}}$.

Now we consider a compositional structure of $\mathcal{X}$ induced by the set $[N] := \{1, \dots, N\}$. Given a subset $A \subseteq [N]$, we define

$$\mathcal{X}_A := \{0, 1\}^A,$$

and the natural projection

$$X_A : \mathcal{X}_{[N]} \to \mathcal{X}_A, \qquad (x_i)_{i \in [N]} \mapsto (x_i)_{i \in A} .$$

In the following, we will abbreviate $\mathcal{X} := \mathcal{X}_{[N]}$. One can view $\overline{\mathcal{P}}(\mathcal{X})$ as the set of joint probability distributions of the binary random variables $\{X_i : i \in [N]\}$. We now use the compositional structure of $\mathcal{X}$ in order to define exponential families in $\mathcal{P}(\mathcal{X})$ given by interaction spaces. Now, decompose $x \in \mathcal{X}$ in the form $x = (x_A, x_{[N]\setminus A})$ with $x_A \in \mathcal{X}_A$, $x_{[N]\setminus A} \in \mathcal{X}_{[N]\setminus A}$, and define $\mathcal{I}_A$ to be the subspace of functions that do not depend on the configurations $x_{[N]\setminus A}$:

$$\mathcal{I}_A \quad := \quad \Big\{ f \in \mathbb{R}^{\mathcal{X}} \ : \ f(x_A, x_{[N]\setminus A}) = f(x_A, x'_{[N]\setminus A})$$

$$\text{for all } x_A \in \mathcal{X}_A, \text{ and all } x_{[N]\setminus A}, x'_{[N]\setminus A} \in \mathcal{X}_{[N]\setminus A} \Big\} .$$

In the following, we apply these interaction spaces as building blocks for more general interaction spaces and associated exponential families [DS83]. The definition of a hierarchical model is based on the notion of a hypergraph [Lau96]:

Definition 2. A *pre-hypergraph* $\mathcal{A}$ is a non-empty subset of $2^{[N]} \setminus \{\emptyset\}$ that contains all atoms $\{i\}$ for $i \in [N]$.

A *hypergraph* is a pre-hypergraph that is (inclusion) complete in the following sense: If $A \in \mathcal{A}$ and $\emptyset \neq B \subseteq A$ it follows that $B \in \mathcal{A}$.

Remark. For technical convenience, we have defined hypergraphs to be complete. In this way, it is easy to define a hierarchical model for each hypergraph. However, the notion of a pre-hypergraph turns out to be more natural in the context of the polytopes and linear codes that we consider below.

Given a hypergraph, we define the associated interaction space by

$$\mathcal{I}_{\mathcal{A}} := \sum_{A \in \mathcal{A}} \mathcal{I}_A.$$

Note that, since a function that depends only on its arguments in A, only depends on its arguments in $B \supseteq A$, it suffices to consider the inclusion maximal elements in $\mathcal{A}$. We denote them by $\mathcal{A}^m$ and have

$$\mathcal{I}_{\mathcal{A}} = \sum_{A \in \mathcal{A}^m} \mathcal{I}_A.$$

We consider the corresponding exponential family:

Definition 3. The *hierarchical model* assigned to the hypergraph $\mathcal{A}$ is the exponential family

$$\mathcal{E}_{\mathcal{A}} := \exp(\mathcal{I}_{\mathcal{A}}).$$

We give two examples for hypergraphs:

Example 4.
(1) Graphical models: Let $G = (V, E)$ be an undirected graph, and define

$$\mathcal{A}_G := \{\emptyset \neq C \subseteq V : C \text{ is a clique with respect to } G\}.$$

Here, a *clique* is a set C that satisfies the following property:

$$a, b \in C, \; a \neq b \quad \Rightarrow \quad \{a, b\} \in E.$$

The exponential family $\mathcal{E}_{\mathcal{A}_G}$ is characterized by Markov properties with respect to G (see [Lau96]).
(2) Interaction order: The hypergraph associated with a given interaction order $k \in \{1, 2, \ldots, N\}$ is defined as

$$\mathcal{A}_{k,N} := \{\emptyset \neq A \subseteq [N] : |A| \leq k\}.$$

If appropriate, we will sometimes drop the N and write $\mathcal{A}_k$. We have defined a corresponding hierarchy of exponential families studied in [Ama01, AK06]:

$$\mathcal{E}_{\mathcal{A}_1} \subseteq \mathcal{E}_{\mathcal{A}_2} \subseteq \cdots \subseteq \mathcal{E}_{\mathcal{A}_N} = \mathcal{P}(\mathcal{X}).$$

The elements of this hierarchy have nice interpretations. It can be seen that the closure of the family $\mathcal{E}_{\mathcal{A}_1}$ contains exactly all probability distributions that factor. This means that

$$p \in \overline{\mathcal{E}_{\mathcal{A}_1}} \Leftrightarrow p = \prod_{i \in [N]} p_i(x_i),$$

where $p_i(x_i)$ are the marginal distributions of p. Generally, an element $p \in \mathcal{E}_k$ will allow a factorization as

$$p = \prod_{A \subseteq [N], |A| = k} \phi_A(x_A),$$

where ϕ_A depends only on x_A. However, the ϕ are not necessarily probability distributions and not unique. Note that $p \in \overline{\mathcal{E}_k} \setminus \mathcal{E}_k, k \geq 2$, does not necessarily admit such a product structure.

We will clarify these definitions in the following simple

Example 5. Consider the case $N = 2$. The configuration space is given as

$$\mathcal{X} := \{0, 1\}^2 = \{(0,0), (0,1), (1,0), (1,1)\}.$$

The vector space of real valued functions $\mathbb{R}^{\mathcal{X}}$ is 4-dimensional and the probability measures form a 3-dimensional tetrahedron. Considering the hypergraphs of fixed interaction order and their exponential families, one has only two examples here: $\mathcal{E}_{\mathcal{A}_{1,2}}$ and $\mathcal{E}_{\mathcal{A}_{2,2}} = \mathcal{P}(\mathcal{X})$, only the first being nontrivial. Figure 1 shows the situation.
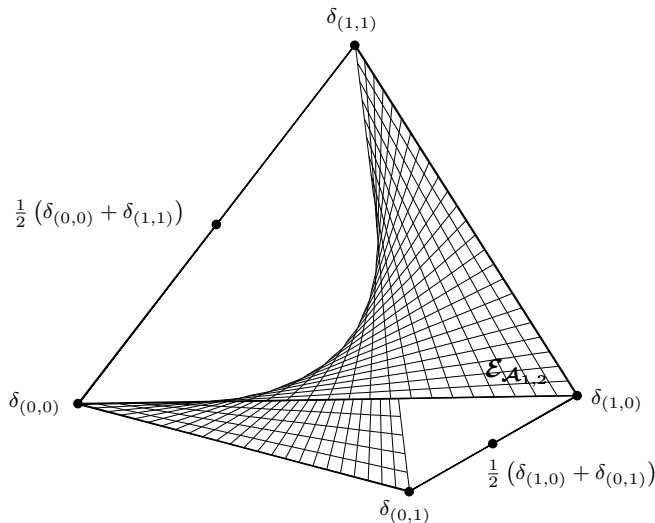

FIGURE 1. The exponential family $\mathcal{E}_{\mathcal{A}_{1,2}}$ in the simplex of probability distributions.

The exponential family $\mathcal{E}_{1,2}$ is a two-dimensional manifold lying inside the simplex. One should already think about this as a square (the two dimensional cube) molded into the simplex.

In the following, we will study the interaction spaces more thoroughly by comparing different generating systems.

3. Generating Systems of Interaction Spaces

In this section, let $\mathcal{A}$ be fixed. In statistics, different representations of exponential families have been considered, each of which has its own benefits and highlights different aspects. We will review a number of these representations. The particular choice of parity functions will allow us to make a link to coding theory.

As we have introduced exponential families, the key concept is the interaction space $\mathcal{I}_{\mathcal{A}}$ which is sometimes also called tangent space to the exponential family. This space completely characterizes the exponential family. However, there is a choice of the parameterization of this space, which has been made differently in different fields. Speaking in terms of linear algebra, one has to choose a generating system of a linear space.

Let $\mathfrak{B} := \{b_k : k \in K\}$ be any finite generating system of $\mathcal{I}_{\mathcal{A}}$. Each such choice gives a different parameterization of the exponential family and a different sufficient

statistics. The parameterization is identifiable if $\mathfrak{B}$ is a basis. The exponential family is parameterized as

$$\mathcal{E}_{\mathcal{A},\mathfrak{B}} = \left\{ p \in \mathcal{P}(\mathcal{X}) : p(x) = Z_\theta^{-1} \exp\left(\sum_{k \in K} \theta_k b_k(x) \right) : \theta \in \mathbb{R}^d \right\},$$

where again Z_θ is the normalization and $d = |\mathfrak{B}|$ equals the number of parameters. In statistical physics the exponent is commonly called the energy.

To each choice of $\mathfrak{B}$ there is a polytope constructed as follows. Consider the vectors

$$b_K(x) = (b_k(x))_{k \in K}$$

Each such vector has as its components the evaluation of every element in $\mathfrak{B}$ at x. The polytope is

$$P_\mathfrak{B} := \operatorname{conv}\left\{ b_K(x) : x \in \mathcal{X} \right\}.$$

Since $\mathcal{A}$ contains all atoms, it can be seen that the polytope has $|\mathcal{X}|$ vertices and the dimension equals the dimension of the exponential family. By applying some classical theorems from statistics, such as the existence and uniqueness of maximum likelihood estimates [Kul68, Csi75], it can be seen that the points of the polytope are in one to one correspondence with points in the closure of the exponential family. As we have introduced it here, it is clear that the different choices of $\mathfrak{B}$ yield different representation of the same polytope in the sense that they are all affinely equivalent. In particular, they have the same face lattice.

The polytope $P_\mathfrak{B}$ encodes in its face lattice the combinatorial structure of the exponential family in the sense that a knowledge of the face lattice gives precise knowledge about the supports of elements in the closure of the exponential family. However, direct computation is infeasible for real world problems.

In statistical physics, and also for various inference methods it is of interest to compute the free energy, given as the logarithm of the partition function. There, variational principles and the techniques of Legendre transform are applied. In this setting the points in the polytope are then the so called dual parameters. See for instance [WJ03].

We will review a number of choices for $\mathfrak{B}$:
Statistical Physics - Potentials. In statistical physics one considers so called potentials [Win03, Geo88]. A potential is a collection of functions $U_A, A \subseteq N$, where $U_A \in \mathcal{I}_A$ and $U_\emptyset = 0$, such that the energy can be written as a linear combination hereof. Typically one has a distinguished state o called the vacuum. A potential is called *normalized* if $U_A(x) = 0$ as soon $x_i = o_i$ for some $i \in A$. Given a strictly positive distribution, a corresponding normalized potential exists and is unique. In our binary setting, choosing $(0, 0, \ldots, 0)$ as the vacuum state, the normalized potential is given by the functions $U_A = c_A \prod_{i \in A} x_i$, where $c_A \in \mathbb{R}$.

One has $\mathfrak{B} = \left\{ \prod_{i \in A} x_i : A \in \mathcal{A} \right\}$, and a basis of the interaction space is given by $\mathfrak{B}$ together with the constant function $x \mapsto 1$. Expanding a function $H \in \mathbb{R}^\mathcal{X}$ in terms of this basis was called the χ-expansion in the works of Caianiello [Cai86, Cai75].

In the case of pair interactions where the hypergraph is given by $\mathcal{A}_{2,N}$, the polytope $P_\mathfrak{B}$ coincides with the so called *correlation polytope* [DL97]. Extending the terminology to an arbitrary hypergraph $\mathcal{A}$, we call $P_\mathfrak{B}$ the *moment polytope*, as each point in it is the vector of moments of some distribution.

Marginals. One representation of an exponential family is given via the linear map that computes the marginals. Denote $\mathcal{A}^m$ the set of inclusion maximal sets in $\mathcal{A}$. Consider the linear map

$$\pi_{\mathcal{A}} : \mathbb{R}^{\mathcal{X}} \to \bigoplus_{A \in \mathcal{A}^m} \mathbb{R}^{\mathcal{X}_A}$$

$$u \mapsto (u_A)_{A \in \mathcal{A}}.$$

That, for a given vector u computes the set of its maximal marginals defined as

$$u_A(x_A) := \sum_{y : X_A(y) = x_A} u(y).$$

When represented as a matrix with respect to the canonical basis, $\pi_{\mathcal{A}}$ has rows indexed by pairs (A, y_A) of a set $A \in \mathcal{A}^m$ and a configuration $y_A \in \mathcal{X}_A$. The columns are indexed by configurations $x \in \mathcal{X}$. Each component then contains the value of the indicator $\mathbb{1}_{\{X_A = y_A\}}$.

Denote the x-th column of this matrix as π_x then, the exponential family is parameterized as

$$p(x) = Z_{\theta}^{-1} \exp(\langle \theta, \pi_x \rangle), \qquad \theta \in \mathbb{R}^d.$$

In terms of these vectors, the polytope is commonly called the marginal polytope. It is represented as 0/1 polytope embedded in a high dimensional space.

An orthogonal basis of characters. In the binary case $\mathcal{X} = \{0, 1\}^{[N]}$, a natural basis for $\mathbb{R}^{\mathcal{X}}$ is given by the characters of $\mathcal{X}$. Here, we assume pointwise addition modulo 2 as the group operation. For every subset $A \in \mathcal{A}$ define the function $e_A : \mathcal{X} \to \{-1, 1\}$ by

$$e_A(x) := (-1)^{E(A, x)}$$

where $E(A, x) := |\{i \in A : x_i = 1\}|$. It can be seen that, if $\mathcal{A}$ is a hypergraph, $\{e_A : A \in \mathcal{A}\}$ together with the constant function $e_{\emptyset} : x \mapsto 1$ is an orthogonal basis of the interaction space $\mathcal{I}_{\mathcal{A}}$. This approach was followed in [KA06]. Various people, starting with Caianiello [Cai75] have called this the η-expansion. Note that if one considers random variables taking values in $\{\pm 1\}$ this basis equals the monomial basis $\{\prod_{i \in A} x_i : A \subseteq [N]\}$ considered above.

A basis of parity functions. Finally, we will introduce yet another basis of $\mathcal{I}_{\mathcal{A}}$ which is derived from the basis of characters. To each $\emptyset \neq A \subseteq [N]$, we define a vector in $\mathbb{R}^{\mathcal{X}}$.

(1) $$f_A(x) := \begin{cases} 1 & \text{if } |\text{supp}(x) \cap A| \text{ is odd} \\ 0 & \text{otherwise.} \end{cases}$$

The following proposition is easily checked:

Proposition 6. *Let* $\mathbb{1} : \mathcal{X} \to \mathbb{R}$ *be the constant function* $x \mapsto 1$. *The set* $\{f_A : A \in \mathcal{A}\} \cup \{\mathbb{1}\}$ *is a basis of* $\mathcal{I}_{\mathcal{A}}$.

One crucial point about choosing this representation is that it gives, if the constant function is omitted, a full dimensional 0/1 polytope, the vertices of which form an additive group and thereby a linear code (see Proposition 10). For all other choices of $\mathfrak{B}$ discussed in this section the image of b_K is not a subgroup of $\{0, 1\}^d$ or the multiplicative group $\{\pm 1\}^d$.

While in the construction of a hierarchical model we assumed a hypergraph, the following polytope is an interesting object of study also in the general case of a pre-hypergraph:

Definition 7. Let $\mathcal{A}$ be a pre-hypergraph. We define

$$\mathcal{F}_{\mathcal{A}} := \text{conv} \left\{ f_{\mathcal{A}}(x) : x \in \mathcal{X} \right\}.$$

If $\mathcal{A}$ is a hypergraph, then this is affinely equivalent to the marginal polytope of the corresponding exponential family. In the case of the hypergraphs $\mathcal{A}_{k,N}$ we write $\mathcal{F}_{k,N} := \mathcal{F}_{\mathcal{A}_{k,N}}$. The rest of the paper is devoted to the study of this class of polytopes.

Remark (CUT-Polytopes). There is a well known [DL97] affine equivalence between CUT polytopes of graphs [Zie00] and binary marginal polytopes:

Namely, to each graph G we can associate the hypergraph $\mathcal{A}_G = V(G) \cup E(G)$. This is distinct from what was called a graphical model above, as not the cliques are considered. Some authors refer to the corresponding statistical model as a *graph model*. From G construct the coned graph $\hat{G}$ with an additional vertex:

$$V(\hat{G}) := V(G) \cup \{*\},$$

and edges

$$E(\hat{G}) := E(G) \cup \{(v, *) : v \in V(G)\}.$$

Then, denoting the CUT polytope of $\hat{G}$ as $CUT(\hat{G})$ one has

$$\mathcal{F}_{\mathcal{A}_G} = CUT(\hat{G}).$$

Using the representation in terms of the vectors $f_{\mathcal{A}_G}(x)$, $x \in \mathcal{X}$, the proof of this equivalence becomes a simple renaming of coordinates.

Remark (Covariance Mapping). As remarked above, in the representation with monomials $\prod_{i \in A} x_i$ one finds the correlation polytope $COR(N)$ as a special case. From the last remark it follows that the CUT-polytope of the complete graph K_{N+1} is equal to $\mathcal{F}_{\mathcal{A}_{2,N}}$. There exists an affine equivalence between $COR(N)$ and $CUT(K_{N+1})$ called the *covariance mapping* [DL97]. It can be seen that this mapping generalizes to a mapping between binary marginal polytopes and the corresponding moment polytopes. It therefore might be suitable to consider the parity representations $\mathcal{F}_{\mathcal{A}}$ of binary marginal polytopes for a generalization of CUT-polytopes to arbitrary (pre)-hypergraphs.

3.1. **Computations and elementary properties.** Using the geometry software `polymake` [GJ00], one can compute linear descriptions of polytopes. As an example, we give here the F-Vectors of $\mathcal{F}_{k,N}$ for the cases $N = 3, 4$. For $N = 5$, the F-Vector is too complicated to be computed by the brute force approach of `polymake`. However, waiting sufficiently long, one can get the 6800 facet defining inequalities of $\mathcal{F}_{3,5}$ and the 3835488 facets of $\mathcal{F}_{4,6}$.

Example 8. In Tables 1 and 2, we give the F-Vectors of $\mathcal{F}_{k,N}$ for $N = 3, 4$, computed using `polymake`. The rows label the dimension of the faces, the columns the value of k. The reader might wonder about the fact that the face lattices of $\mathcal{F}_{k,N}$ are up to a certain dimension isomorphic the face lattice of the simplex. This property, commonly called neighborliness, follows from a general result in [Kah08]. The last row refers to whether the polytope is simple or not.

$d\backslash k$	1	2	3
0	8	8	8
1	12	28	28
2	6	56	56
3	1	68	70
4	-	48	56
5	-	16	28
6	-	1	8
7	-	-	1
sum	27	225	255
simple	y	n	y

TABLE 1. Face structure of $\mathcal{F}_{k,3}$

$d\backslash k$	1	2	3	4
0	16	16	16	16
1	32	120	120	120
2	24	560	560	560
3	8	1780	1820	1820
4	1	3872	4368	4368
5	-	5592	8008	8008
6	-	5060	11440	11440
7	-	2600	12868	12870
8	-	640	11424	11440
9	-	56	7952	8008
10	-	1	4256	4368
11	-	-	1680	1820
12	-	-	448	560
13	-	-	64	120
14	-	-	1	16
15	-	-	-	1
sum	81	20297	65025	65535
simple	y	n	n	y

TABLE 2. Face structure of $\mathcal{F}_{k,4}$

In the following, we will list elementary properties of $\mathcal{F}_{k,N}$ that follow easily from the definition.

(i) $\mathcal{F}_{1,N}$ is the N-cube.
(ii) $\mathcal{F}_{N,N}$ is the $(2^N - 1)$-dimensional simplex.
(iii) every $\mathcal{F}_{k,N}$ has dimension $d = |\mathcal{A}_{k,N}|$.
(iv) every $\mathcal{F}_{k,N}$ has 2^N vertices.
(v) $(0, \ldots, 0)$ is a vertex.
(vi) every $\mathcal{F}_{k,N}$ is a projection of the $(2^N - 1)$-dimensional simplex $\mathcal{F}_{N,N}$ along coordinate axes.
(vii) For every $\mathcal{F}_{k,N}$, there is a projection along coordinate axes that projects it to the N-cube $\mathcal{F}_{1,N}$.

Remark. In [HS02] it was remarked that $\mathcal{F}_{N-1,N}$ has exactly 4^{N-1} facets. The extreme points of these facets are also known. A set $\mathcal{Y} \subsetneq \mathcal{X}$ defines a face if and only if it contains neither $\mathcal{U} := \left\{ x \in \mathcal{X} : f_{[N]}(x) = 1 \right\}$ nor its complement. Note that the set $\mathcal{U}$ and its complement are exactly the set of configurations with a fixed parity. As the vertices of $\mathcal{F}_{N-1,N}$ have only one affine dependency, it is not difficult to prove this fact using the Gale transform. By the above $\mathcal{F}_{N-1,N}$ is combinatorially isomorphic to the so called cyclic polytope [Zie94].

In the following, we develop the connection to coding theory.

4. A Link to Coding Theory

We briefly recall the definition of a linear code. For a detailed introduction into coding theory see for instance [van99]. Consider the finite field $\mathbb{F}_2 = (\{0,1\}, \oplus, \odot)$ with addition and multiplication mod 2. In coding theory, one studies particularly vector spaces over this field.

Definition 9. A binary $[n,k]$-*linear code* is a linear subspace L of $\mathbb{F}_2^n$ such that $\dim L = k$. A *generator matrix* G for L is a k by n matrix which has as its rows a basis of L. Given L one can find an equivalent[1] code such that it has a generator matrix in *standard form,* i.e. $G = (E_k, H)$, where E_k is the k by k identity matrix.

The following proposition states that the vertices of $\mathcal{F}_\mathcal{A}$ form a linear code for any pre-hypergraph $\mathcal{A}$. A special case of this connection has been mentioned in Example 2 in [WJ03].

Proposition 10. *Let $\{0,1\}^{\mathcal{A}}$ be considered as a vector space over the finite field $\mathbb{F}_2$. Then the image of $\mathcal{X}$ under $f_\mathcal{A}$ is a linear subspace. If we also consider $\mathcal{X} = \mathbb{F}_2^N$ as a vector space over $\mathbb{F}_2$, $f_\mathcal{A}$ is an injective homomorphism between vector spaces. Its image forms an $[|\mathcal{A}|, N]$-linear code. A generator matrix in standard form has as its rows the vectors $f_\mathcal{A}(e_i)$ for $i = 1, \ldots, N$, where e_i is the i-th unit vector in $\mathbb{F}_2^N$.*

Proof. Since scalar multiplication is trivial, we only need to show

$$(2) \qquad f_\mathcal{A}(x \oplus y) = f_\mathcal{A}(x) \oplus f_\mathcal{A}(y) \qquad \text{for } x, y \in \mathcal{X}.$$

Let $A \in \mathcal{A}$, it suffices to show the identity for f_A. To do so, introduce

$$M := \{i \in A : (x_i = 1 \wedge y_i = 0) \vee (x_i = 0 \wedge y_i = 1)\},$$
$$M_x := \{i \in A : x_i = 1\},$$
$$M_y := \{i \in A : y_i = 1\}.$$

Then $f_A(x \oplus y) = |M|$, $f_A(x) = |M_x|$, and $f_A(y) = |M_y|$. We find that M is the symmetric difference of M_x and M_y:

$$M = M_x \triangle M_y.$$

Since $|M_x \triangle M_y| = |M_x| + |M_y| - 2 |M_x \cap M_y|$, we have that in $\mathbb{F}_2$

$$|M| = |M_x| \oplus |M_y|,$$

and therefore (2) holds. We now show that $f_\mathcal{A}$ is injective. To see this, assume that $f_\mathcal{A}(x) = f_\mathcal{A}(y)$. Since $\mathcal{A}$ contains all atoms $\{i\} \subseteq [N]$, we get for every $i \in [N]$:

---

[1] Two codes are called equivalent if one can be transformed into the other by applying a permutation on the positions in the codewords, and for each position a permutation of the symbols.

$f_{\{i\}}(x) = f_{\{i\}}(y)$. This implies $x_i = y_i$ and, hence, $x = y$. Since $\mathcal{X}$ considered as an $\mathbb{F}_2$ vector space has dimension N, also $f_{\mathcal{A}}(\mathcal{X})$ has dimension N and therefore forms an $[\mathcal{A}, N]$-linear code. $\square$

Remark. To write down the generator matrix, one has to impose a numbering on the elements in $\mathcal{A}$. If the numbering is in such a way that $A_i = \{i\}$ for $1 \leq i \leq N$, then the generator matrix is in standard form.

Remark. An important property of a linear code is its distance, which is defined as the minimal Hamming distance between different elements of the code. For the hierarchical model of the hypergraph $\mathcal{A}_{k,N}$, the distance of the code is given by

$$d = \sum_{l=0}^{k-1} \binom{N-1}{l}.$$

Proof. Let $d(x, y)$ denote the hamming distance of $x, y \in \mathcal{X}$. If $d(x, y) = 1$, then $d(f_{\mathcal{A}_k}(x), f_{\mathcal{A}_k}(y))$ equals the number of subsets of $[N]$ which contain a given element and have cardinality at most k. $\square$

In the following, we will elaborate on the opposite direction. Let $2^N \geq s \geq N$. Assume we are given an $[s, N]$ linear code. Without loss of generality, we assume that it has a generator matrix in standard form. We will construct a pre-hypergraph $\mathcal{A}$ from the columns of the generator matrix. Since $\mathcal{A}$ is a set, while the columns are a list, repetitions of columns will be lost. If one considers only non-repetitive codes, then our construction is injective, and the codewords are given by the vertices of $\mathcal{F}_{\mathcal{A}}$.

Let $E_N \in \mathbb{F}_2^{N \times N}$ denote the identity matrix in dimension N. Assume the generator matrix $G = (E_N, H) \in \mathbb{F}_2^{N \times s}$ has no 2 identical columns. (This implies $s \leq 2^N$.) Denote by $\{e_i : i = 1, \ldots, N\}$ the canonical basis of $\mathbb{F}_2^N$. Using the columns of H, we define sets

$$A_j := \{i \in [N] : H_{ij} = 1\}, \qquad j = 1, \ldots, s - N$$

and then,

$$\mathcal{A} := \{\{1\}, \ldots \{N\}, A_1, \ldots, A_{s-N}\}.$$

Note that the elements of $\mathcal{A}$ are numbered in a natural way such that we can use $\mathcal{A}$ as an index set for the columns of $G = (G_{i,A})_{i=1,\ldots,N, \, A=\{1\},\ldots,\{N\},A_1,\ldots,A_{s-N}}$.

To see that $\{f_{\mathcal{A}}(e_i) : i = 1, \ldots, N\}$ is the set of rows of the generator matrix, we evaluate

$$f_A(e_i) = \delta_{\{i \in A\}} = G_{i,A}$$

which holds by definition of the A_j.

Summarizing, every binary linear code (in standard form) corresponds to a pre-hypergraph. However, two codes that differ only in repetitions of columns in the generator matrix will be mapped to the same pre-hypergraph. Then, if it is a hypergraph, the linear code is the linear code of a hierarchical model.

5. CLASSIFICATION

As we have seen, the polytopes $\mathcal{F}_{\mathcal{A}}$ are full dimensional polytopes such that the vertices form a linear code. In this last section, we classify all polytopes with this property. Then we investigate which of them can be realized as polytopes of

hierarchical models. For a convex polytope P, let $V(P)$ denote the vertex set of P. For $n \in \mathbb{N}$, put

$$C_n := [0,1]^n,$$
$$W_n := \{0,1\}^n = V(C_n).$$

Hence $(W_n, \oplus)$ is an Abelian group that is canonically isomorphic to $(\mathbb{F}_2^n, \oplus)$. We consider W_n as a subset of $\mathbb{R}^n$ and write "$\oplus$" whenever we mean addition modulo 2, while "$+$" means ordinary addition in $\mathbb{R}^n$.

In the following, we develop an algorithm that determines - by induction for every $n \in \mathbb{N}$ - all polytopes $P \subseteq \mathbb{R}^n$ with $V(P) \subseteq W_n$ satisfying the following conditions:

(I) $(V(P), \oplus)$ is a subgroup of $(W_n, \oplus)$.
(II) P has dimension n.

Note that the number of vertices of such a polytope is a power of two. Of course, the full n-dimensional cube $P = C_n$ satisfies (I) and (II). To start the induction, we remark that there are no further such polytopes in the cases $n = 1$ and $n = 2$. For $n = 3$, the 3-dimensional regular simplex S with

$$V(S) = \{(0,0,0), (1,1,0), (1,0,1), (0,1,1)\}$$

satisfies (I) and (II), too.

More generally, by [Wen06, Theorem 2.2], we have the following

Proposition 11. *For $n \geq 3$, the following statements are equivalent:*

(1) *$(W_n, \oplus)$ contains some subgroup U such that $\mathrm{conv}(U)$ is a regular simplex of dimension n.*
(2) *$n + 1$ is some power of 2.*

In the case $n = 3$, the full 3-cube as well as the regular simplex mentioned above are the only polytopes satisfying conditions (I) and (II). Note that also

$$\{(0,0,0), (1,0,0), (0,1,1), (1,1,1)\}$$

determines a subgroup of $(W_3, \oplus)$; however, the convex closure has dimension 2.

For fixed $n \geq 2$, define the bijections $\pi_0 : \mathbb{R}^n \times \{0\} \to \mathbb{R}^n$ and $\pi_1 : \mathbb{R}^n \times \{1\} \to \mathbb{R}^n$ by

$$\pi_0(x_1, \ldots, x_n, 0) := (x_1, \ldots, x_n),$$
$$\pi_1(x_1, \ldots, x_n, 1) := (x_1, \ldots, x_n).$$

For $1 \leq i \leq n$ put

$$H_i := \{(x_1, \ldots, x_n) \in \mathbb{R}^n : x_i = 0\},$$
$$H_i' := \left\{(x_1, \ldots, x_n) \in \mathbb{R}^n : x_i = \frac{1}{2}\right\}.$$

Moreover, let $z := \left(\frac{1}{2}, \ldots, \frac{1}{2}\right)$ denote the center of the n-cube C_n.

To determine recursively all 0/1-polytopes $P \subseteq \mathbb{R}^n$ that fulfill (I) and (II), we prove first the following

Proposition 12. *Suppose that $n \geq 2$ and that $P \subseteq \mathbb{R}^n$ is a 0/1-polytope satisfying (I) and (II). Assume that $(U, \oplus)$ is a subgroup of $(V, \oplus) := (V(P), \oplus)$ with $|V : U| = 2$. Then the following statements are equivalent:*

(i) The polytope $Q \subseteq \mathbb{R}^{n+1}$, given by

$$(3) \qquad V(Q) = \pi_0^{-1}(U) \cup \pi_1^{-1}(V \setminus U),$$

has dimension $n + 1$.

(ii) There **does not** exist some index i with $1 \leq i \leq n$ such that $U \subseteq H_i$. In other words, none of the affine hyperplanes H_i' separates $\operatorname{conv}(U)$ and $\operatorname{conv}(V \setminus U)$.

(iii) One has $z \in \operatorname{conv}(U) \cap \operatorname{conv}(V \setminus U)$.

(iv) One has $\operatorname{conv}(U) \cap \operatorname{conv}(V \setminus U) \neq \emptyset$.

Proof. **(i)** $\rightarrow$ **(ii):**

Suppose that $U \subseteq H_i$ holds for some i with $1 \leq i \leq n$. Put

$$\tilde{H} := \left\{ (x_1, \ldots, x_n, x_{n+1}) \in \mathbb{R}^{n+1} : x_{n+1} = x_i \right\}.$$

Since $P = \operatorname{conv}(V)$ has dimension n and since $|V : U| = 2$, we must have $x_i = 1$ whenever $(x_1, \ldots, x_n) \in V \setminus U$. This means that $V(Q)$ - and hence also Q - is contained in the n-dimensional hyperplane $\tilde{H}$, in contradiction to (i).

(ii) $\rightarrow$ **(iii):**

For $1 \leq i \leq n$, let $\alpha_i : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$ denote the linear map given by $\alpha_i(x_1 \ldots, x_n) := x_i$. By assumption, $\alpha_{i \upharpoonright U}$ is surjective for $1 \leq i \leq n$. Hence we have

$$|\{u \in U : \alpha_i(u) = 0\}| = |\{u \in U : \alpha_i(u) = 1\}| \quad \text{for } 1 \leq i \leq n.$$

This means that

$$z = \frac{1}{|U|} \sum_{u \in U} u \in \operatorname{conv}(U),$$

where the sum is taken in $\mathbb{R}^n$.

Now fix $v_1 \in V \setminus U$. Since $V \setminus U = \{v_1 \oplus u : u \in U\}$, we get also

$$|\{v \in V \setminus U : \alpha_i(v) = 0\}| = |\{v \in V \setminus U : \alpha_i(v) = 1\}| \quad \text{for } 1 \leq i \leq n,$$

and hence

$$z = \frac{1}{|V \setminus U|} \sum_{v \in V \setminus U} v \in \operatorname{conv}(V \setminus U).$$

(iii) $\rightarrow$ **(iv)** is trivial.

(iv) $\rightarrow$ **(i):**

Consider the projection $\pi : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^n$ given by

$$\pi(x_1, \ldots, x_n, x_{n+1}) = (x_1, \ldots, x_n).$$

Suppose that the assertion is wrong; hence Q is contained in some n-dimensional -homogeneous- hyperplane $G \subseteq \mathbb{R}^{n+1}$. Since

$$\pi(V(Q)) = U \dot{\cup} (V \setminus U) = V,$$

the polytope Q has the same dimension as $P = \operatorname{conv}(V)$, that is n. Thus, the restriction $\pi_{\upharpoonright G}$ is a linear isomorphism from G onto $\mathbb{R}^n$, and there exists some $\mathbb{R}$-linear map $\alpha : \mathbb{R}^n \rightarrow G$ satisfying

$$(\alpha \circ \pi)(w) = w \qquad \text{for all } w \in G.$$

By definition of $V(Q)$, this means:

$$\alpha(U) = \pi_0^{-1}(U),$$
$$\alpha(V \setminus U) = \pi_1^{-1}(V \setminus U).$$

Hence, $\alpha(\mathrm{conv}(U)) = \mathrm{conv}(\alpha(U))$ and $\alpha(\mathrm{conv}(V \setminus U)) = \mathrm{conv}(\alpha(V \setminus U))$ are linearly separated by the affine hyperplane

$$K := \left\{ (x_1, \ldots, x_n, x_{n+1}) \in \mathbb{R}^{n+1} : x_{n+1} = \frac{1}{2} \right\}.$$

By (iv) this is impossible.   □

Example 13. We investigate the statement of Proposition 12 for polytopes corresponding to a pre-hypergraph $\mathcal{A}$. We start by considering the matrix which has as its rows the vectors $f_{\mathcal{A}}(x)$, where $\mathcal{A} = 2^{[N]} \setminus \{\emptyset\}$. The rows are labeled by the binary strings of length N, that is by $\mathcal{X}$, while the columns are indexed by the non-empty subsets of $[N]$. Therefore the rows of this matrix are the coordinates of the vertices of the simplex $\mathcal{F}_{N,N}$:

x	$\{1\}$	$\cdots$	$\{N\}$	$\{1,2\}$	$\cdots$	$\{1,2,3\}$	$\cdots$	$[N]$
$(00\ldots 0)$	0	$\cdots$	0	0	$\cdots$	0	$\cdots$	0
$(00\ldots 1)$	0	$\cdots$	1	$f_{\{1,2\}}(x)$	$\cdots$	$f_{\{1,2,3\}}(x)$	$\cdots$	$f_{[N]}(x)$
$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$	$\vdots$
$(11\ldots 1)$	1	$\cdots$	1	$f_{\{1,2\}}(x)$	$\cdots$	$f_{\{1,2,3\}}(x)$	$\cdots$	$f_{[N]}(x)$

We note the following facts:

- The columns of this matrix are exactly the $2^N - 1$ non-zero binary strings of length N.
- There are $2^N - 1$ subgroups U of index 2 of the N-cube, which correspond to the columns of the matrix. To define them let a column A be fixed, then put $U := \{x \in \mathcal{X} : f_A(x) = 0\}$. The maps $f_A : \mathcal{X} \to \{0,1\}$ are exactly the $2^N - 1$ surjective homomorphisms having the nontrivial subgroups as their kernels.
- The vertices of every polytope $\mathcal{F}_{\mathcal{A}}$ are given by deleting columns from this matrix that correspond to sets not in $\mathcal{A}$.
- In particular, by restriction to the first N columns, we get the vertices of the N-cube $\mathcal{F}_{1,N}$.

Now, assume that P is the N-cube. We choose a column of the matrix, corresponding to a subgroup of index 2. There are two possibilities. If we choose a column corresponding to an atom, then (ii) is wrong, the dimension does not grow when adding this column to the coordinates (as we have doubled a coordinate). If, on the other hand, we choose a column corresponding to a set A with cardinality two or more, then we are in the situation of Proposition 12, since (ii) holds. The lift (3) will be full dimensional, and its vertices are given by the submatrix with columns $\{1\}, \ldots, \{N\}, A$. Continuing from here, choosing another subgroup, the dimension will grow if and only if it does not correspond to one of the sets $\{1\}, \ldots, \{N\}, A$. Iteratively, the choices narrow down and, finally, when all columns have been chosen, the polytope Q is a simplex.

We will now formalize this procedure. For a fixed polytope P as in Proposition 12, put

$$U_i := V(P) \cap H_i \qquad \text{for } 1 \le i \le n.$$

Clearly, conditions (I) and (II) imply that $|V(P) : U_i| = 2$ holds whenever $1 \le i \le n$.

Based on the equivalence of (i) and (ii) in Proposition 12, we are now able to prove that the following algorithm yields recursively all 0/1-polytopes satisfying (I) and (II).

Algorithm 14.
Initialization for $n = 1$*:*

  - $\mathfrak{P}_1 := \{[0, 1]\}$.

Step $n \to n + 1$*: Based on* $\mathfrak{P}_n$ *construct a new set* $\mathfrak{P}_{n+1}$ *consisting of all polytopes* Q *such that there exists* $P \in \mathfrak{P}_n$ *with*

  - $Q = P \times [0, 1]$ *or*
  - $Q \subseteq \mathbb{R}^{n+1}$ *with*

(4) $$V(Q) = \pi_0^{-1}(U) \cup \pi_1^{-1}(V(P) \setminus U),$$

   where U *runs through all subgroups of* $(V(P), \oplus)$ *with* $|V(P) : U| = 2$ *and* $U \neq U_i$ *for* $1 \leq i \leq n$.

Remark. Note that in the case $Q = P \times [0, 1]$, the number of vertices is doubled, while in the other cases the number of vertices of Q equals the number of vertices of P. Furthermore, it is interesting to see that the two possible operations commute in the following sense. Starting from some cube W_n, lifting it to W_{n+1} and then choosing a subgroup U to apply the lift (4) gives the same polytope as choosing the subgroup $\pi(U)$ from W_n and then taking the prism over the lifted polytope, where $\pi : \mathbb{R}^{n+1} \to \mathbb{R}^n$ is the canonical projection. Therefore, all polytopes that are constructed by the algorithm can be thought of as lifted cubes W_n.

The classification will be complete with:

Theorem 15. *For all* $n \in \mathbb{N}$*, the set* $\mathfrak{P}_n$ *in Algorithm 14 consists of all* n*-dimensional 0/1 polytopes that satisfy conditions (I) and (II).*

Proof. First we show that all polytopes $Q \in \mathfrak{P}_{n+1}$ satisfy conditions (I) and (II), with n replaced by $n + 1$. This is clear in the case of the prism $Q = P \times [0, 1]$.

If Q satisfies (4), then clearly $(V(Q), \oplus)$ is a subgroup of $(W_{n+1}, \oplus)$, because U is a subgroup of $(V(P), \oplus)$ with $|V(P) : U| = 2$. Moreover, (ii) $\to$ (i) in Proposition 12 implies that Q has dimension $n + 1$, because $U \neq U_i$ for $1 \leq i \leq n$. Hence, Q satisfies conditions (I) and (II).

Vice versa, assume that $Q \subseteq \mathbb{R}^{n+1}$ fulfills (I) and (II). Consider again the projection $\pi : \mathbb{R}^{n+1} \to \mathbb{R}^n$ onto the first n coordinates, and put $P := \pi(Q)$. Since Q has dimension $n + 1$, P has dimension n. If $\pi_{|V(Q)}$ is not injective, then Q is the prism $P \times [0, 1]$, because $(V(Q), \oplus)$ is a subgroup of $(W_{n+1}, \oplus)$. If $\pi_{|V(Q)}$ is injective, put

$$U := \{(x_1, \ldots, x_n) \in V(P) | (x_1, \ldots, x_n, 0) \in Q\}.$$

Then $(U, \oplus)$ is a subgroup of $(V(P), \oplus)$ with $|V(P) : U| = 2$, because Q has dimension $n + 1$. Moreover, equation (4) holds for U as just defined. Finally, Proposition 12, (i) $\to$ (ii) shows that $U \neq U_i$ for $1 \leq i \leq n$. Hence, our algorithm includes the determination of Q. $\square$

As a first application of Theorem 15 we can count the number of n-dimensional polytopes that satisfy conditions (I) and (II). Let $c_n := |\mathfrak{P}_n|$. For $1 \leq k \leq n$, let

$c_n(k)$ denote the number of all 0/1 polytopes $P \subseteq \mathbb{R}^n$ with $|V(P)| = 2^k$ that satisfy (I) and (II). Then one has obviously

$$(5) \qquad c_n = \sum_{k=1}^{n} c_n(k).$$

We have $c_n(k) = 0$ for $2^k \leq n$, because a polytope with at most n vertices cannot have dimension n. Furthermore, we have clearly $c_n(n) = 1$ for all $n \in \mathbb{N}$.

As mentioned already in Example 13, a 0/1-polytope that satisfies (I), (II), and $|V(P)| = 2^k$ has among its vertices exactly $2^k - 1$ subgroups of index 2. Hence by ignoring the groups $U_i = V(P) \cap H_i$ for $1 \leq i \leq n$, we get

Corollary. *For $k \leq n < 2^k$ one has*

$$c_{n+1}(k) = c_n(k-1) + c_n(k)(2^k - n - 1).$$

The first few values are given in the Table 3. It is easy to compute this number

n \ k	1	2	3	4	5	6	7	8	c_n
1	1								1
2	0	1							1
3	0	1	1						2
4	0	0	5	1					6
5	0	0	15	16	1				32
6	0	0	30	175	42	1			248
7	0	0	30	1605	1225	99	1		2960
8	0	0	0	12870	31005	6769	219	1	50864

TABLE 3. The number of n-dimensional 0/1 polytopes with 2^k vertices that form a group.

also for larger values of n. For instance

$$c_{28} = 718897730072178204358180468879825453986397667929112558174208$$

$$c_{100} \approx 2.77 \cdot 10^{644}$$

Finally, using the Corollary we can show that, among the full dimensional 0/1-polytopes with 2^k vertices the convex hulls of linear codes are exceptional. For $1 \leq k \leq n$, let $d_n(k)$ denote the number of all 0/1 polytopes with 2^k vertices satisfying only condition (II). Hence, the number d_n of all 0/1 polytopes of dimension n trivially satisfies

$$(6) \qquad d_n \geq \sum_{k=1}^{n} d_n(k).$$

Moreover, we get

Proposition 16.   *(i) For $4 \leq n < 2^k < 2^n$, one has*

$$(7) \qquad d_n(k) \geq 2^k(2^n - 2^k)c_n(k) > n2^{n-1}c_n(k).$$

(ii) We have

$$\lim_{n \to \infty} \frac{c_n}{d_n} = 0.$$

Proof.   (i) Suppose that U is a proper subgroup of $(W_n, \oplus)$ with $\dim(\mathrm{conv}(U)) = n$ and $|U| = 2^k$.

If U' is another subgroup of $(W_n, \oplus)$ with $|U'| = |U|$, then we have

$$|U \cap U'| \leq 2^{k-1} < 2^{n-1}$$

and, hence,

$$|U \setminus U'| \geq 2^{k-1} > \frac{n}{2} \geq 2.$$

This means

(8) $$|U \setminus U'| \geq 3.$$

There are $2^k(2^n - 2^k)$ subsets V of W_n with $|V| = 2^k$ and $|V \setminus U| = |U \setminus V| = 1$; namely, these are all sets of the form

(9) $$V = (U \setminus \{u_0\}) \cup \{v_0\} \quad \text{with } u_0 \in U, v_0 \in W_n \setminus U.$$

For V as in (9), we get $\dim(\mathrm{conv}(V)) = n$, because otherwise, $U \setminus \{u_0\}$ would be contained in a -unique- hyperplane H with $v_0 \in H$, a contradiction to $v_0 \notin U$. Together with (8), we obtain the first inequality in (7). The second one is trivial in view of $2^k \leq 2^{n-1}$.

(ii) By (5), (6), and (7) we get for $n \geq 4$:

$$\frac{c_n}{d_n} \leq 2\frac{c_n - 1}{d_n - 1}$$

$$\leq 2\left(\sum_{k=1}^{n-1} c_n(k)\right)\left(\sum_{k=1}^{n-1} d_n(k)\right)^{-1}$$

$$\leq 2(n2^{n-1})^{-1} = \frac{2^{2-n}}{n}.$$

This proves the second statement.   □

As a concluding remark, we study the question of constructing a statistical model from a given polytope. Assume $P \subseteq \mathbb{R}^n$ satisfies (I) and (II), when does it come from a hierarchical model? To begin with, observe that the number m of vertices of P is a power of 2, since it must divide the number of vertices of the cube W_n. We write $m = 2^N$. By Theorem 15, we know that P can be constructed using the algorithm. It is constructed from the N-cube by applying several steps of the second type in Algorithm 14. Therefore, every such polytope corresponds to a subset of columns $\{1\}, \ldots, \{N\}, A_{N+1}, \ldots, A_s$ in the matrix of coordinates, or equivalently to a pre-hypergraph. However, this pre-hypergraph is not unique. If we are given only a list of vertices, then there are several possibilities to choose a generator matrix in standard form. As an example consider the polytope $\mathcal{F}_{2,3}$:

x	$\{1\}$	$\{2\}$	$\{3\}$	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$
$(0,0,0)$	0	0	0	0	0	0
$(0,0,1)$	0	0	1	0	1	1
$(0,1,0)$	0	1	0	1	0	1
$(0,1,1)$	0	1	1	1	1	0
$(1,0,0)$	1	0	0	1	1	0
$(1,0,1)$	1	0	1	1	0	1
$(1,1,0)$	1	1	0	0	1	1
$(1,1,1)$	1	1	1	0	0	0

In fact every 3 by 3 submatrix which is, after permuting rows and columns, the identity matrix gives a generator matrix in standard form. Obviously, there are several such choices. Here, for instance we can choose the canonical basis corresponding to the atoms:

x	$\{1\}$	$\{2\}$	$\{3\}$	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$
$(1,0,0)$	1	0	0	1	1	0
$(0,1,0)$	0	1	0	1	0	1
$(0,0,1)$	0	0	1	0	1	1

On the other hand, we can also reorder the columns and choose

x	$\{3\}$	$\{12\}$	$\{13\}$	$\{1\}$	$\{2\}$	$\{2,3\}$
$(1,1,1)$	1	0	0	1	1	0
$(0,1,0)$	0	1	0	0	1	1
$(1,1,0)$	0	0	1	1	1	1

When the generator matrix is chosen, one can apply the method given in Section 4 to construct a pre-hypergraph. For our first generator matrix we get back the hypergraph we started with, for the second choice we read of the pre-hypergraph

$$\mathcal{A}' := \{\{1'\}, \{2'\}, \{3'\}, \{1', 3'\}, \{1', 2', 3'\}, \{2', 3'\}\},$$

where we have introduced new units $\{1', 2', 3'\}$ corresponding to the first three columns. This shows the ambiguity due to the choice of a particular generator matrix if only the code is given.

Acknowledgment

References

[AK06] N. Ay and A. Knauf. Maximizing multi-information. *Kybernetika*, 42:517–538, 2006.

[Ama01] S. Amari. Information geometry on hierarchy of probability distributions. *IEEE Transactions on Information Theory*, 47:1701–1711, 2001.

[Cai75] E. Caianiello. Synthesis of boolean nets and time behaviour of a general mathematical neuron. *Biol. Cybernetics*, 18, 1975.

[Cai86] E. Caianiello. Neuronic equations revisited and completely solved. In G. Palm and A. Aertsen, editors, *Brain Theory*. Springer, Berlin, 1986.

[Csi75] Imre Csiszár. *I*-divergence geometry of probability distributions and minimization problems. *Ann. Probab.*, 3:146–158, 1975.

[DL97] M.M. Deza and Monique Laurent. *Geometry of Cuts and Metrics*. Algorithms and Combinatorics. Springer, 1997.

[DS83] J. N. Darroch and T. P. Speed. Additive and multiplicative models and interactions. *The Annals of Statistics*, 11(3):724–738, Sep 1983.

[FWK03] Jon Feldman, Martin Wainwright, and David R. Karger. Using linear programming to decode linear codes. In *Conference on Information Sciences and Systems*, 2003.

[Geo88] H.-O. Georgii. *Gibbs Measures and Phase Transitions*. Number 9 in de Gruyter Studies in Mathematics. de Gruyter, 1988.

[GJ00] E. Gawrilow and M. Joswig. `polymake`: a framework for analyzing convex polytopes. In *Polytopes—combinatorics and computation*, volume 29, pages 43–47. Birkhäuser, 2000.

[HS02] S. Hoşten and S. Sullivant. Gröbner bases and polyhedral geometry of reducible and cyclic models. *Journal of Combinatorial Theory: Series A*, 100(2):277–301, 2002.

[KA06] T. Kahle and N. Ay. Support sets of distributions with given interaction structure. In *Proceedings of the WUPES 2006*, 2006.

[Kah08] Thomas Kahle. Neighborliness of marginal polytopes. 2008. submitted, arXiv:0809.0786.

[Kul68] Solomon Kullback. *Information Theory and Statistics*. Dover, New York, 1968.

[Lau96] Steffen L. Lauritzen. *Graphical Models*. Oxford Statistical Science Series. Oxford University Press, 1996.

[van99] J.H. van Lint. *Introduction to Coding Theory*. GTM. Springer, third edition edition, 1999.

[Wen06] W. Wenzel. Regular simplices inscribed into the cube and exhibiting a group structure. *Journal of Combinatorial Mathematics and Combinatorial Computing*, 59:213–220, 2006.

[Win03] Gerhard Winkler. *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods*. Springer, 2nd edition, 2003.

[WJ03] Martin J. Wainwright and Michael I. Jordan. Variational inference in graphical models: The view from the marginal polytope. In *Allerton Conference on Communication, Control, and Computing*, 2003.

[Zie94] Günter M. Ziegler. *Lectures on Polytopes*, volume 152 of *GTM*. Springer Verlag, 1994.

[Zie00] Günter M. Ziegler. Lectures on 0/1-polytopes. In G. Kalai and Günter M. Ziegler, editors, *Polytopes - Combinatorics and Computation*, pages 1–41. Birkhäuser, 2000.

Max Planck Institute for Mathematics in the Science, Inselstrasse 22, D-04103 Leipzig, Germany

E-mail address: kahle@mis.mpg.de

Fachhochschule Magdeburg, Fachbereich Wasser- und Kreislaufwirtschaft, Breitscheidstrasse 2, D-39114 Magdeburg, Germany

E-mail address: walter@math.uni-bielefeld.de

Max Planck Institute for Mathematics in the Science, Inselstrasse 22, D-04103 Leipzig, Germany, and Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

E-mail address: nay@mis.mpg.de